

基于小样本学习的高光谱遥感图像分类算法

张 婧 袁细国

(西安电子科技大学 计算机科学与技术学院, 陕西 西安 710071)

摘 要 高光谱图像(Hyperspectral Image, HSI)具有丰富的地物空间信息和光谱信息,为地物的精细分类提供了有利条件.当前高光谱遥感图像分类方法,在标记样本较少的情况下,其分类性能相对有限.那么,如何充分发挥小样本学习领域的相关理论与方法,以提升高光谱遥感图像分类性能,是一个意义重要且具有挑战性的问题.为此,本文以标记的小样本为背景,对现有基于深度学习的高光谱遥感图像分类算法进行改进,其核心思路在于:利用基于深度三维神经网络的嵌入模型从训练数据集中学习投影函数,进而将视觉特征映射到语义表示,同时通过最小化损失函数使其类内间距较小、类间间距较大,再使用训练好的网络嵌入模型将每个未标记的目标类表示在相同的嵌入空间中,最后,分别选用欧氏距离和卷积神经网络作为距离度量方式,计算嵌入空间中嵌入与类原型之间的距离,从而实现投影后的样本分类.本文在印第安松、萨利纳斯和帕维亚大学数据集上进行了实验,其结果表明改进的算法具有较好的高光谱遥感图像分类性能.

关键词 小样本学习;高光谱遥感图像;高光谱遥感图像分类;深度学习

中图分类号 N39

文献标识码 A

0 引言

高光谱成像光谱仪捕获地球表面同一区域的几十到数百窄波段连续光谱图像,相当于获得每个像素对应的近乎连续的光谱曲线.高光谱图像(Hyperspectral Image, HSI)具有高光谱分辨率及丰富的光谱空间信息,为准确鉴别不同土地覆盖类别提供了优势条件.因此,高光谱图像分类被充分应用于精准农业、环境保护、军事应用等诸多领域,并发挥了重要的作用,现有卫星传感器如 Hyperion、Hy-Map 和 AVIRIS 都是高光谱图像数据的良好来源.

近年来,高光谱遥感图像分类算法也得到了深入的研究.高光谱遥感图像分类是根据图像对应特征将每个像素分配到一个特定的类中.一般来说,有监督光谱分类器,诸如支持向量机^[1]、随机森林^[2],通常比无监督分类器的分类性能更好.然而,由于在高光谱图像中,收集标记样本通常是耗时耗力的,高维特征空间与有限数据样本之间的矛盾仍然是有监督分类器的瓶颈所在.虽然特征提取^[3,4]和波段选择^[5]有效缓解了高光谱图像数据标记样本少以及特征维度高等特性导致的休斯现象,深度学习^[6-9]、半监督学习^[10-12]、自监督学习^[13]在高光谱图像分类领域也取得了不错的表现,高光谱图像分类算法仍然很难利用有限的标记样本得到期望的分类性能.

为了解决这一问题,Dopido 等人^[14]提出了一个自学习框架,选择最有用的未标记样本并使用分类器预测其伪标签,然后将未标记样本转为标记样本加入到训练数据集中,但是这一方法会不断累积并传播分类器可能出现的分类错误.李伟等人^[15]提出了像素对方法以显著增加训练样本的数目,潘斌等人^[16]提出了一个简化的深度学习模型 MugNet 以减少对标记样本的需求,詹等人^[17]提出了一个基于一维生成对抗网络的半

收稿日期:2020-05-20

基金项目:国家自然科学基金面上项目(61571341)资助

通讯作者:袁细国,男,汉族,博士,副教授,研究方向:大数据分析与人机智能,计算生物信息学,图像分析等;E-mail: xiguoyuan@mail.xidian.edu.cn.

监督框架来生成与训练集中标记样本相似的伪样本以增加标记样本的数目. 这些方法尽管已经取得了很大的进展,但仍然需要相当数量的标记样本,仅用少量的标记样本对高光谱图像进行分类仍然存在一定的困难. 因此,如何将小样本学习领域的显著进展拓展应用于高光谱遥感图像分类领域极具意义.

人类非常擅长在仅见过极少量的样本,甚至从未见过样本的情况下识别一个新物体,在人类的这一能力的启发下,许多学者期望机器学习模型能够在一定类别的大量数据上学习如何学习,从而也仅需要新类别的极少量样本就能快速识别新类,即小样本学习. 最近许多方法在小样本学习领域取得了显著进展. 基于初始化的方法是学习如何更新参数,其中基于模型初始化方法利用先验知识来约束假设空间的复杂性,使得假设空间要小得多,如 MANN^[18] 使用外部存储器来存储样本表现和类标签值的绑定信息,学习良好的模型初始化,利用有限标记样本通过少量更新步骤得到新类分类器;基于优化器的方法利用先验知识来寻找假设空间中最优假设的参数,如 MAML^[19] 不断由元学习器根据学习器的性能更新参数. 基于距离度量学习的方法是学习如何进行比较,如匹配网络^[20] 利用余弦相似度来度量训练集嵌入和查询集嵌入之间的距离从而预测查询集类别,原型网络^[21] 利用欧氏距离来度量查询集与各类别的类原型之间的距离从而进行分类,而关系网络^[22] 则利用卷积神经网络来对度量方式进行建模. 基于数据增强的方法则利用先验知识来扩充训练数据集,如从训练数据集中每个样本转换为多个具有变化的样本来增强训练集^[23-25] 或利用更大的相似数据集的输入-输出对来增强训练集^[26].

刘冰等人已经将小样本学习领域的成果应用于高光谱图像分类领域,先后提出了 S-CNN 算法^[27] 和 DFSL 算法^[28],前者基于孪生网络提取能够更好地表示输入样本对间相似性的特征,后者在训练数据集中学习一个度量空间并将其泛化到测试数据集的类中,但这些方法无法通过端到端的方式训练,均需要在提取特征的基础上使用 SVM 分类器来预测类标签.

为此,本文以小样本为背景,针对现有基于深度学习的高光谱遥感图像分类算法提出了两种改进算法. 本文将每个样本表示为该像素及其空间邻域内所有像素以联合光谱和空间信息,因此输入到网络中时,每个样本为一个三维数据立方体. 算法大致可以分为三步,首先在训练数据集中学习嵌入模型,进而将视觉特征映射到语义表示,同时通过最小化损失函数使其类内间距较小、类间间距较大,当训练完成后得到的嵌入模型可以视为一个投影函数;然后使用投影函数将每个未标记的目标类表示在相同的嵌入空间中;最后,分别选用欧氏距离和卷积神经网络作为距离度量方式,计算嵌入空间中嵌入与类原型之间的距离,从而实现投影后的样本分类.

1 基于小样本学习的高光谱遥感图像分类算法

本文以小样本标签数据为背景,针对现有的基于小样本学习的高光谱图像分类算法存在的问题对其进行改进,构建了两种算法,即基于深度学习的高光谱图像小样本分类算法和基于关系网络的高光谱图像小样本分类算法,下面对每一种算法的理论和实现过程进行详细描述.

1.1 基于深度学习的 HSI 小样本分类算法

为了能够在测试数据集上进行小样本分类,模型必须利用在训练数据集上习得的知识来学习识别测试数据集中新的类别的能力. 现存的域迁移方法很难直接应用于小样本分类任务,因为在测试数据集上仅有极少量标记样本可用. 因此,问题的关键在于学习领域不变的、可泛化的特征表示和/或可用于测试数据集的分类模型.

在小样本分类任务中,关键在于使用中间层语义表示来使得源域数据与目标域数据在语义上相关,而这些语义表示通常编码在高维向量空间中. 常见的语义表示包括语义属性和语义词向量. 给定一个预定义的语义表示,各类均能表示为一个属性向量或一个语义词向量,即类原型. 语义表示通常是通用且共享的,可以用来在训练数据集和测试数据集之间传递知识,从而识别新的、不可见的类. 使用嵌入模型从训练数据集中学习投影函数,投影函数将视觉特征映射为语义表示. 利用学习到的投影函数将不可见类的未标记样本表示在同一个嵌入空间. 然后通过计算嵌入空间中嵌入与类原型的相似度使用分类模型对每个投影后的样本进行分类.

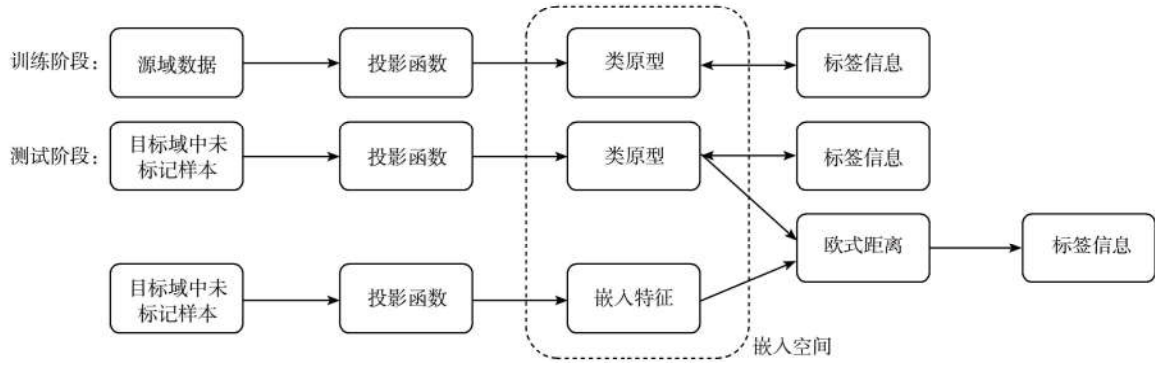


图 1 基于深度学习的 HSI 小样本分类算法结构示意图

受此启发,基于深度学习的 HSI 小样本分类算法大致可分为三个阶段.首先,使用基于深度残差三维卷积神经网络的嵌入模型从训练数据集中学习到一个投影函数,学习到的投影函数能够将视觉特征映射到可泛化特征,在训练过程中通过最小化损失函数令投影函数得到的嵌入特征类内间距较小且类间间距较大,且这个投影函数将对训练中没有出现的类也执行相同的操作.其次,利用预先训练好的网络提取测试集中每个样本的特征.最后,通过计算嵌入空间中嵌入与类原型之间的欧氏距离使用最近邻分类器对每个投影后的样本进行分类.之所以算法采用简单的最近邻分类器来避免搜索额外参数,是因为仅有少数标记样本可用,而常用的分类器(如 SVM)需要使用交叉验证来确定最优参数.基于深度学习的 HSI 小样本分类算法结构如图 1 所示.

1.1.1 嵌入模型及训练策略.由于深度 CNN 模型具有很强的非线性表达能力,同时考虑到残差学习可以更好地训练深度三维卷积神经网络(D-3-D CNN),本文最终依然采用深度残差三维卷积神经网络(D-Res-3-D CNN)对度量空间进行参数化.这个度量空间可以被视为一个投影函数.算法利用这一具有可学习参数的投影函数来计算每个类的原型.

为了使习得的投影函数适用于小样本分类场景,算法使用抽样小批量(episode)来模拟小样本任务.从训练数据集中随机抽取一个类子集形成一个训练批次来计算梯度并更新网络参数.对于每一

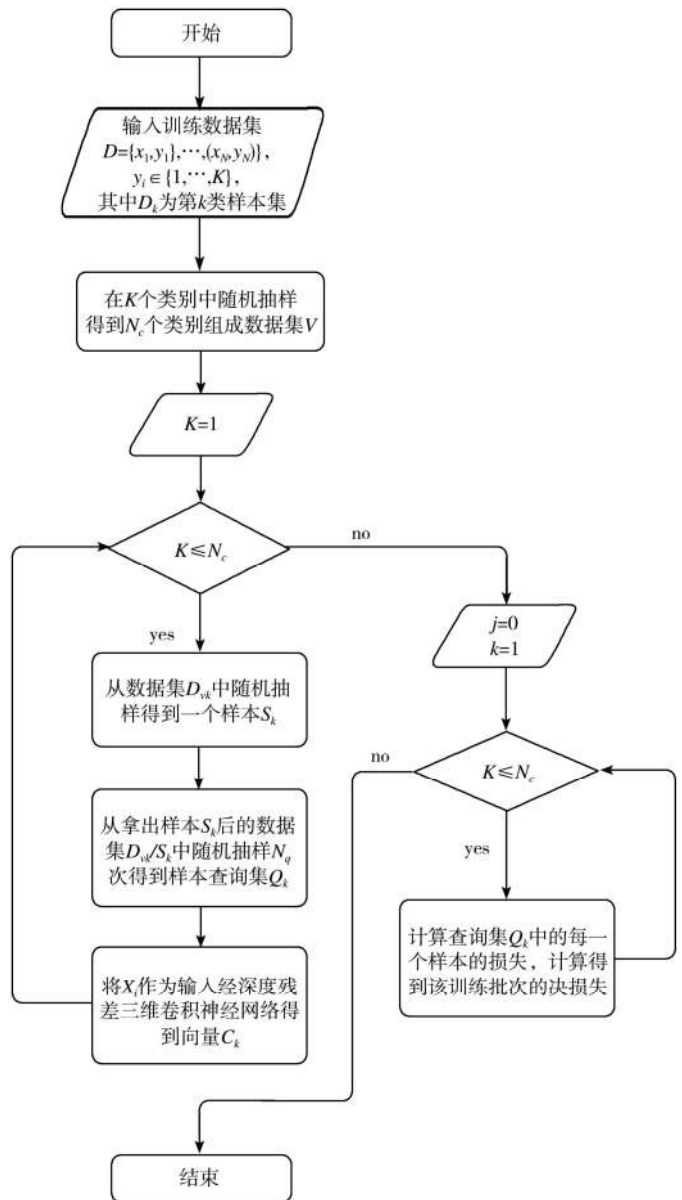


图 2 训练过程

训练批次,从每一个类中随机抽取一个样本作为支持集以模拟测试数据集上进行的小样本分类任务,剩余样本的子集作为查询集.将支持集和查询集的样本输入到网络中,通过网络提取其嵌入特征.对于查询集的每一个样本,其类分布函数基于嵌入空间中与支持集样本的距离由式(1)^[28]计算得到

$$p_{\varphi}(y = k | x) = \frac{e^{-d(f_{\varphi}(x), c_k)}}{\sum_{k=1}^{N_c} e^{-d(f_{\varphi}(x), c_k)}} \quad (1)$$

其中 c_k 表示支持集第 k 类的深层特征, x 表示查询集样本的嵌入特征, y 表示 x 的类标签, $d(\cdot)$ 为欧氏距离函数. 损失函数如式(2)^[28]所示, 最小化损失函数会使得查询集样本的嵌入特征 $f_\varphi(x)$ 与该样本对应类 k 类原型 c_k 之间的距离尽可能小, 而与各个类原型之间的距离尽可能大.

$$J(\varphi) = -\log p_\varphi(y = k | x) = d(f_\varphi(x), c_k) + \log \sum_{k=1}^{N_c} e^{-d(f_\varphi(x), c_k)}. \quad (2)$$

以一个训练批次为例, 训练过程的程序流程图如图 2, 以便网络能够学习到一个投影函数, 该投影函数得到的嵌入特征具有类内间距较小且类间间距较大的特点.

嵌入模型中选择 Mish 函数作为激活函数, $f = x \times \tanh(\ln(1 + e^x))$, 其函数图像见图 3^[29]. 之所以选择换用 Mish 函数作为激活函数的原因是: 其一, Mish 函数以上无边界, 即正值可以达到任何高度, 避免了由于封顶而导致的饱和; 其二, 不同于 ReLU 函数的硬零边界, Mish 函数对负值的轻微允许理论上能够允许更好的梯度流; 其三, Mish 函数相比 ReLU 函数更为平滑, 因此能够允许更好的信息深入神经网络, 从而得到更好的准确性和泛化.

由于陈等^[30]实验证明先批量归一化再激活的常见做法会呈现出锯齿形的优化行为, 阻止网络参数沿梯度方向进行更新. 而先激活再进行批量归一化则会使得整个训练过程更加稳定, 特别是在前 30 个训练批次中能够获得更好的泛化性能. 此外, 将批量归一化层置于激活层之后, 网络还得到了更快的收敛速度和更好的泛化性能.

如图 4 所示, 网络包含两个残差块, 两个最大池化层以及一个三维卷积层. 三维卷积层下方数字表示该层卷积核的数目及卷积核的尺寸; 最大池化层表示三维最大池化, 卷积核大小设置为 $2 \times 2 \times 4$; 选择 Mish 函数作为激活函数 $f(x) = x \tanh(\ln(1 + e^x))$. 残差块中不使用池化层. 残差块与三维最大池化层相连, 以减少计算和聚集特征. 光谱维数的步长仍然设置为 4, 空间维数的步长设置为 2. 最后, 将特征映射平铺成一维向量. 利用图 2 所示的策略对网络进行训练, 从而使得网络学习到一个投影函数能使得嵌入特征具有类内间距小且类间间距大的特点.

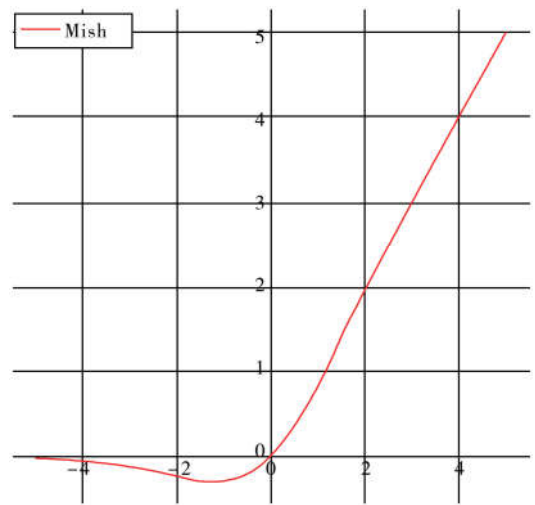


图 3 Mish 激活函数

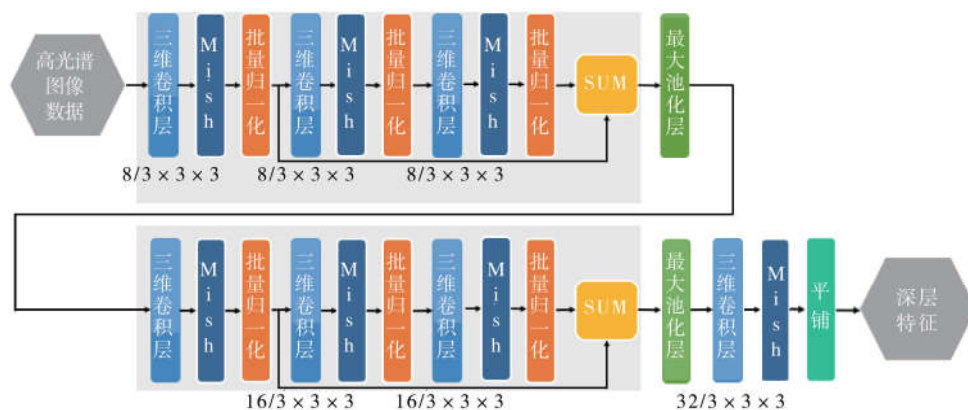


图 4 网络模型结构

Li 等人已证明 $3 \times 3 \times 3$ 的卷积核是高光谱图像分类中三维卷积神经网络的有效选择, 因此三维卷积核的大小设置为 $3 \times 3 \times 3$. 其它网络参数见表 1.

1.1.2 分类策略. 如图 5 所示, 测试数据集分类主要包含三个步骤: 首先, 通过预先训练好的深度残差三维 CNN 提取嵌入特征, 将每个未标记的目标类表示在相同的嵌入空间中; 其次, 计算嵌入空间中嵌入与类原型的欧氏距离; 最后, 使用最近邻分类模型对每个投影后的样本进行分类. 为了验证该算法在仅有少量标记样本的情况下的分类性能, 将测试样本的预测标签生成的分类图与真值图进行比较.

表 1 D-Res-3D-CNN 相关参数

层	层输出	卷积核
输入	9 × 9 × 100	N/A
卷积层 1	9 × 9 × 100 × 8	3 × 3 × 3 × 8
卷积层 2	9 × 9 × 100 × 8	3 × 3 × 3 × 8
卷积层 3	9 × 9 × 100 × 8	3 × 3 × 3 × 8
短路径		卷积层 1+卷积层 3
最大池化层 1	5 × 5 × 25 × 8	2 × 2 × 4
卷积层 4	5 × 5 × 25 × 16	3 × 3 × 3 × 16
卷积层 5	5 × 5 × 25 × 16	3 × 3 × 3 × 16
卷积层 6	5 × 5 × 25 × 16	3 × 3 × 3 × 16
短路径		卷积层 4+卷积层 6
最大池化层 2	3 × 3 × 7 × 16	2 × 2 × 4
卷积层 7	1 × 1 × 5 × 32	3 × 3 × 3 × 32
平铺层	160	N/A
输出层	160	N/A

1.2 基于关系网络的 HSI 小样本分类算法

基于关系网络的 HSI 小样本分类算法结构示意图如图 6 所示. 分类算法分为两个模块:特征提取模块和关系度量模块. 特征提取模块首先使用嵌入模型从训练集中学习到一个投影函数,然后利用投影函数得到测试集中每个样本的嵌入特征. 相比基于深度学习的 HSI 小样本分类算法对于未分类样本的分类依据仅仅是单一且固定的距离度量方式,即比较该样本嵌入后的特征向量与各类原型之间的欧氏距离,基于关系网络的 HSI 小样本分类算法引入了关系网络来对度量方式进行建模,通过训练一个卷积神经网络来学习距离的度量方式. 关系度量模块通过卷积神经网络比较嵌入空间中嵌入与类原型之间的相似度得到关系得分,同属一类的样本对的关系得分高,不同属一类的样本对的关系得分低.

具体来说,样本为高光谱遥感图像数据中的每个像素对应的数据立方体. 支持集中的样本经过投影函数映射到嵌入空间中的嵌入特征,各类别的类原型是嵌入空间中该类支持集样本的均值向量. 查询集中的样本经过投影函数映射到嵌入空间中的嵌入特征. 然后将嵌入特征与每个类别对应的类原型在深度维上拼接. 模型训练时的损失函数为均方误差(MSE),其计算公式为

$$J(\varphi) = \sum_{i=1}^{N_c \times N_s} \sum_{j=1}^{N_c \times N_q} (r_{i,j} - 1 \cdot (y_i == y_j))^2, \tag{3}$$

其中 N_c 为训练数据集中类别数, N_s 为支持集中的每类样本数, N_q 为查询集中的每类样本数, $r_{i,j}$ 为样本 x_i 和样本 x_j 的关系得分, y_i, y_j 表示样本 x_i 和样本 x_j 的标签值.

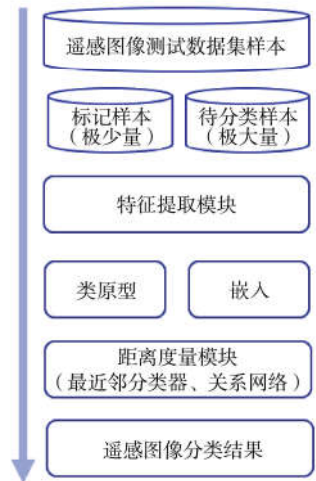


图 5 测试数据集分类流程图

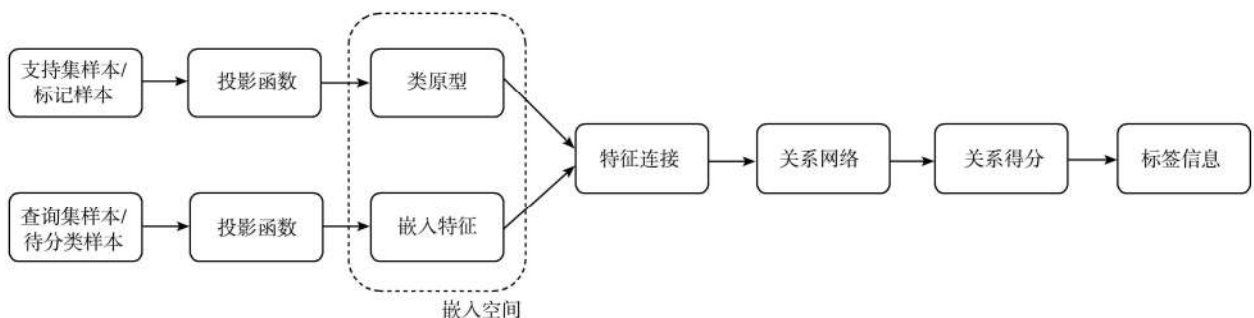


图 6 基于关系网络的 HSI 小样本分类算法结构示意图

1.2.1 嵌入模型及训练策略.如图7所示,嵌入模型与图4基本上一致,但最后对各个通道得到的特征不再进行平铺处理,而是在深度维进行拼接,即将32个 $3 \times 3 \times 7$ 的特征在深度维上拼接为 $3 \times 3 \times 224$ 的特征.利用图2所示的策略对网络进行训练,从而使得网络学习到一个投影函数能使得嵌入特征具有类内间距小且类间间距大的特点.

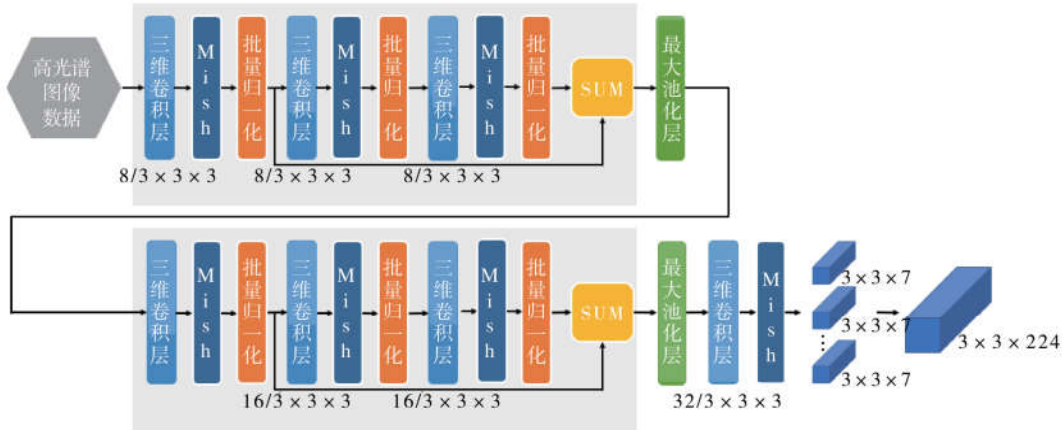


图7 网络模型结构

1.2.2 关系网络.如图8所示,关系网络由两个二维卷积块和两个全连接层组成,卷积核大小为 $3 \times 3 \times 3$,使用批量归一化和ReLU激活函数,步长为 2×2 的最大池化层.除最后一层全连接层使用Sigmoid函数,其他全连接层的激活函数均为ReLU函数.此外,由于关系网络更多地关注关系度量,相比二元分类问题更像是一种回归,因此本文将损失函数换为均方误差.



图8 关系网络结构

训练时,首先在查询集和支持集中随机抽取样本输入到网络中得到深层特征向量,如果支持集中每个训练类的样本数不为0,则对该类所有样本的特征向量进行逐元素求和,然后将类级特征映射与查询集样本的特征向量拼接在一起,再输入到关系网络中通过式(4)得到关系分数,每个分数对应该样本归属各类的概率

$$R_{i,j} = g_{\varphi}(C(f_{\varphi}(x_i), f_{\varphi}(x_j))), i = 1, 2, 3, \dots, n, \quad (4)$$

其中 $f_{\varphi}(\cdot)$ 表示投影函数, $C(\cdot)$ 表示拼接操作, $g_{\varphi}(\cdot)$ 表示关系网络.

1.2.3 分类策略.如图9所示,对测试集中待分类样本进行分类时,首先通过预先训练好的深度残差三维CNN提取嵌入特征,将每个未标记的目标类表示在相同的嵌入空间中;然后,利用预训练好的关系网络计算嵌入特征向量与类原型之间的关系分数从而得到分类结果.

2 实验结果与分析

2.1 实验数据集简介

采用Chikusei数据集^[31]、KSC数据集和Botswana数据集来训练深度三维卷积神经网络.本文采用印第安松数据集、萨利纳斯数据集和帕维亚大学数据集来测试节4.2所提算法的分类性能.

Chikusei数据集是由Headwall Hyperspec-VNIR-C成像传感器在日本茨城町Chikusei的农业和城市地区拍摄,光谱波长范围为 $0.363-1.018 \mu\text{m}$,空间分辨率为 2.5 m .该高光谱数据集由128个光谱波段组成.该图像尺寸为 2517×2335 ,共5877195个像素,其中仅包含77592个地物像素,这些像素被划分为19个类,其假彩色图和分类真值图如图10所示.



图9 关系网络测试数据集分类流程图

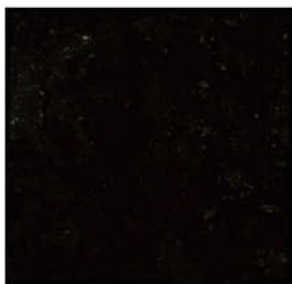


图 10 Chikusei 图像

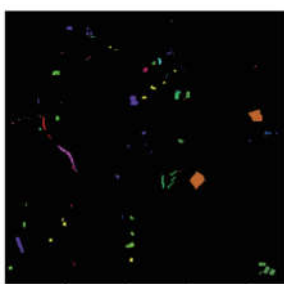
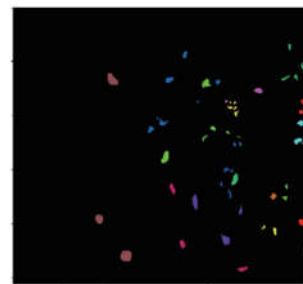


图 11 KSC 图像



KSC 数据集是由 AVIRIS 在佛罗里达州肯尼迪航天中心上空大约 20 km 的高度获取,光谱波长范围为 0.4-2.5 μm ,空间分辨率约为 18 m. 该数据集由 176 个光谱波段组成. 该图像的尺寸为 512×614 像素,共 314368 个像素,其中仅有 5211 个地物像素,这些像素被划分为 13 个类. 其假彩色图和分类真值图如图 11 所示.

Botswana 数据集是由 NASA EO-1 卫星在博茨瓦纳的奥卡万戈三角洲上空获取的,光谱波长范围为 0.4-2.5 μm ,空间分辨率约为 30 m. 该数据集由 145 个光谱波段组成. 该图像的尺寸为 1476×256 像素,共 377856 个像素,其中仅包含 3248 个地物像素,这些像素被划分为 14 个类,其假彩色图和分类真值图如图 12 所示.

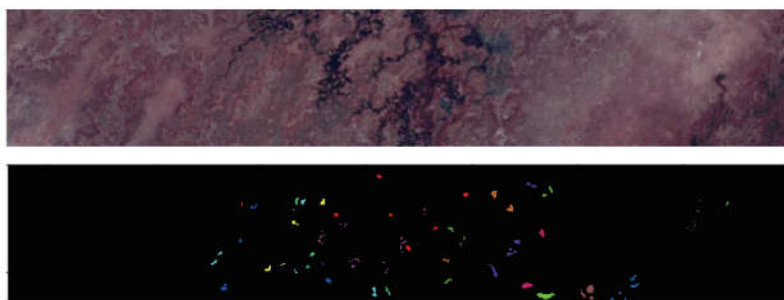


图 12 Botswana 图像

印第安松数据集是由 AVIRIS 在印第安纳州西北部的印第安松试验场上空采集的,光谱波长范围为 0.4-2.5 μm ,空间分辨率约为 20 m. 该数据集由 200 个光谱波段组成. 该图像的尺寸为 145×145 ,因此共有 21025 个像素,其中仅包含 10249 个地物像素,这些像素被划分为 16 个类,其假彩色图和分类真值图如图 13 所示.



图 13 印第安松图像

萨利纳斯数据集是由 AVIRIS 成像光谱仪在加利福尼亚州萨利纳斯山谷上空采集,光谱波长范围为 0.4-2.5 μm ,空间分辨率达到了 3.7 m. 该数据集由 204 个光谱波段组成. 该图像的尺寸为 512×217 ,因此共有 111104 个像素,其中仅包含 54129 个地物像素,这些像素被划分为 16 个类,包括休耕地、芹菜等,其假彩色图和分类真值图如图 14 所示.



图 14 萨利纳斯图像

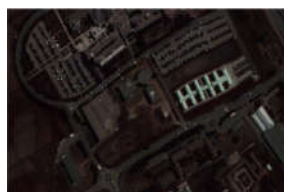


图 15 帕维亚大学图像

帕维亚大学数据集是由 ROSIS 在意大利北部帕维亚上空飞行时捕捉到的两个场景,光谱波长范围为 0.43-0.86 μm ,空间分辨率为 1.3 m. 帕维亚大学数据集由 103 个光谱波段和 610×340 像素组成,该数据集共有 207400 个像素,其中仅包含 42776 个地物像素,这些地物像素被划分为 9 个类,其假彩色图和分类真值

图如图 15 所示.

2.2 算法性能评价指标

对高光谱图像进行分类之后,基于数据的分类真值图以及算法得出的分类结果图,可以评估该分类算法的分类准确性.对于高光谱分类的评价指标主要是总正确率(Overall Accuracy, OA),总正确率(OA)为每个像元的输出标签和其真值标签相符的概率

$$OA = \frac{n_{true}}{n}, \quad (5)$$

其中 n_{true} 是得到正确预测结果的样本数, n 是总测试样本数.

2.3 实验流程及实验结果

训练数据集包含 Chikusei 数据集、KSC 数据集和 Botswana 数据集.这些数据集涵盖了不同的地域和不同的空间分辨率,因此可以保证样本的多样性.该训练数据集共有 46 个类,每个类的样本数不同.为了设计具有相等样本数的训练类,我们实施以下两种情形:(1) 每个训练类的样本个数为 200,训练类共有 40 个;(2) 每个训练类的样本个数为 300,训练类共有 25 个.为了保证深度残差三维 CNN 的输入维数相同,本文采用蔡阳健教授提出的基于图表示的波段选择方法^[5]选择 100 个波段进行训练,各数据集选择的波段编号见表 2 和表 3.本文每个样本的空间邻域大小设置为 9×9 .因此,输入到网络中时,每个样本为 $9 \times 9 \times 100$ 的数据立方体.

表 2 训练数据集选择的波段编号

数据集	选择的波段编号
KSC	155,20,76,0,47,62,18,31,90,139,45,96,60,52,42,32,27,106,36,126,134,75,83,118,161,72,103,48,92,37,104,121,142,169,137,89,162,156,81,57,26,168,86,109,70,165,130,123,43,2,166,4,115,122,133,11,54,82,9,102,53,175,111,167,108,85,35,99,64,12,40,63,117,113,100,98,71,69,97,61,51,73,50,170,146,22,95,107,3,120,15,114,125,124,144,16,132,173,143,84
Botswana	29,24,140,94,108,141,101,71,123,23,111,130,47,69,19,8,110,98,107,78,83,15,86,62,27,118,102,93,9,56,49,129,30,33,42,31,41,109,143,85,137,60,35,44,36,113,112,122,138,64,51,132,59,88,12,99,124,96,106,48,67,66,82,104,54,92,38,84,116,90,95,14,81,70,75,91,128,79,34,17,55,11,10,133,4,3,45,127,58,139,121,114,2,103,136,57,40,74,1,76
Chikusei	108,64,63,55,95,56,113,122,92,99,84,88,106,27,57,36,17,31,67,118,35,124,68,30,104,80,100,3,123,120,44,10,54,5,83,6,112,0,20,109,12,62,7,76,47,90,21,58,34,117,46,25,86,116,125,111,81,42,59,8,28,126,29,19,15,43,60,61,98,85,2,26,13,79,40,33,96,32,51,115,23,93,101,77,121,69,49,1,50,78,66,53,38,82,37,75,9,18,24,74

为了验证所提算法的分类性能,测试训练集包含印第安松数据集、萨利纳斯数据集、帕维亚大学数据集.对于帕维亚大学数据集,本文取其前 100 个波段作为网络输入;对于印第安松和萨利纳斯数据集则取其前 200 个波段作为输入,其中前 100 个波段和后 100 个波段先后经深度残差三维 CNN 网络提取特征后,将得到的两个深层特征连接组成一个新的特征向量用于分类.

训练批次数目设定为 10000,学习率设定为 0.001, N_c 设定为 20, N_q N_9 .此外,卷积核的数量也会影响 D-Res-3-D CNN 的分类性能^[28].卷积核数目太小(如 2,4)会大大降低分类准确率,而卷积核数目太大(如 32)也会降低分类准确率,因此本文第一个残差块的卷积核数设置为 8 和 16.第二个残差块的卷积核数设置为第一个残差块的卷积核数的两倍,即 16 和 32.在 4 个测试数据集上,网络的两个残差块的卷积核数目不同的情况下总正确率(OA)如表 4 所示.由表可见,大部分情况下,第一个残差块的卷积核数目为 16 时的分类性能要优于卷积核数目为 8 时,但是性能提升幅度不大且 16 个卷积核会显著增加训练时间,因此本文将卷积核数目设定为 8.

测试数据集中,每类随机抽取的标记样本数目也会影响分类性能,分类总正确率(OA)随着标记样本数目的增加而增加.每类随机抽取的标记样本数目为 5、10、15、20、25 时,四种算法在测试数据集上分类结果如表 5-表 7.与 Liu 等^[28]提出的 DFSL 算法且分类器分别选用最近邻分类器和支持向量机时相比,更换激活函数为 Mish 函数以及调换归一化和激活顺序后的改进算法在印第安松数据集上取得了更好的表现,而使用关系网络进行关系度量的算法则在萨利纳斯数据集上具有较理想的分类结果.

表 3 测试数据集选择的波段编号

数据集	选择的波段编号
Salinas	6,4,176,96,107,24,123,50,111,160,150,32,182,196,138,39,135,51,166,34,174,144,95,179,191,125,162,90,168,57,143,80,28,119,164,157,199,9,198,5,75,102,40,165,69,109,155,148,87,92,197,200,88,68,31,54,137,16,183,118,64,52,116,70,46,132,19,142,65,37,33,112,147,133,79,184,126,104,189,18,71,190,141,55,193,77,124,29,170,178,115,53,131,99,36,140,15,152,106,45,139,195,93,26,72,14,163,3,128,202,22,41,13,110,74,20,113,153,173,35,49,73,0,85,120,181,76,156,105,100,56,134,1,12,38,48,159,91,27,98,161,11,127,86,167,60,25,117,59,101,146,114,172,23,103,136,187,62,89,63,21,121,43,171,180,151,61,10,7,188,185,108,44,83,58,186,158,130,192,17,129,84,78,94,66,30,145,97,149,154,203,67,175,8,42,82,177,169,81,122
Indian Pines	108,68,36,112,56,85,37,198,76,178,103,101,191,154,60,71,177,195,175,158,66,104,163,114,193,74,39,171,173,32,30,197,65,91,2,43,99,20,45,179,18,132,134,136,143,180,78,44,184,116,139,122,93,149,35,165,109,42,161,3,50,120,164,151,174,152,194,24,26,144,199,185,53,187,90,72,31,82,38,98,97,40,188,19,142,67,159,70,17,57,64,29,79,80,25,8,14,28,61,105,48,11,86,117,7,100,5,0,133,102,62,137,190,160,4,87,166,95,140,10,77,55,121,75,172,27,73,147,69,118,156,170,34,186,131,54,157,21,92,59,153,176,128,63,16,167,168,47,123,115,189,89,135,130,88,146,1,6,15,52,126,148,155,127,141,113,22,129,169,125,49,138,124,119,192,196,9,106,33,110,51,13,181,58,81,12,145,94,46,183,83,23,182,111,107,96,84,162,150,41
Pavia U	82,29,33,37,20,98,53,26,15,4,74,0,34,47,92,99,36,12,79,58,13,2,91,67,18,30,50,96,46,101,21,93,62,73,61,84,56,94,9,95,89,75,38,8,59,69,10,16,76,57,31,54,88,3,43,24,102,39,87,19,68,44,17,5,25,28,35,45,80,48,49,65,77,72,83,27,78,7,100,51,97,90,85,42,6,55,41,70,14,63,11,60,71,32,52,66,22,1,81,86

表 4 四个测试集上 10 次运行平均总正确率(OA)

	8	16
印第安松	66.97 ± 2.61	67.28 ± 2.16
萨利纳斯	87.94 ± 2.23	88.06 ± 1.53
帕维亚大学	77.18 ± 4.11	78.12 ± 3.11

由于 Mish 激活函数更为平滑且对负值轻微允许,嵌入模型使用 Mish 函数理论上能够允许更好的梯度流,得到更好的准确性和泛化性,从而提升分类效果.此外,反转批量归一化和激活的顺序,变为先激活再批量归一化会使得整个训练过程更加稳定,特别是在前 30 个训练批次中能够获得更好的泛化性能,也可以提高在测试数据集上的分类效果.印第安松数据集以及萨利纳斯数据集中,不同的地物具有较为相似的光谱曲线,使用单一且固定的距离度量方式不再满足要求,使用关系网络对度量方式进行建模能够更为灵活地得到嵌入与各类原型之间的距离,从而提升分类性能.

表 5 印第安松数据集上各算法的总正确率(OA)

	L=5	L=10	L=15	L=20	L=25
DFSL+KNN	66.19 ± 2.53	72.64 ± 1.44	77.44 ± 1.42	80.84 ± 1.23	82.94 ± 1.19
DFSL+SVM	66.68 ± 3.42	73.62 ± 5.39	77.55 ± 3.26	81.23 ± 1.72	82.82 ± 2.55
基于深度学习的分类算法	68.61 ± 2.23	75.36 ± 5.01	80.38 ± 7.36	82.24 ± 2.60	84.80 ± 2.57
基于关系网络的分类算法	67.62 ± 4.17	75.02 ± 3.74	78.42 ± 6.57	81.14 ± 9.62	83.04 ± 7.12

表 6 帕维亚大学数据集上各算法的总正确率(OA)

	L=5	L=10	L=15	L=20	L=25
DFSL+KNN	73.72 ± 6.08	81.79 ± 4.91	86.41 ± 2.21	89.77 ± 1.76	90.38 ± 1.26
DFSL+SVM	66.47 ± 4.08	70.45 ± 8.60	85.44 ± 3.92	88.14 ± 6.76	89.99 ± 7.07
基于深度学习的分类算法	75.76 ± 5.50	81.53 ± 4.57	85.77 ± 4.24	92.31 ± 1.94	91.83 ± 2.29
基于关系网络的分类算法	70.85 ± 4.23	80.56 ± 9.67	85.93 ± 6.30	90.00 ± 4.31	91.28 ± 2.71

表 7 萨利纳斯数据集上各算法的总正确率(OA)

	L=5	L=10	L=15	L=20	L=25
DFSL+KNN	85.99 ± 2.59	90.02 ± 0.92	91.46 ± 1.15	92.35 ± 0.68	92.55 ± 0.76
DFSL+SVM	84.58 ± 6.96	86.29 ± 2.71	88.36 ± 3.85	92.62 ± 1.34	91.88 ± 3.54
基于深度学习的分类算法	87.09 ± 3.01	91.21 ± 1.03	92.59 ± 2.42	92.98 ± 0.62	93.05 ± 1.96
基于关系网络的分类算法	87.54 ± 4.61	90.61 ± 2.68	92.14 ± 4.22	93.65 ± 2.56	94.70 ± 2.49

3 结论

深度学习方法、半监督学习方法、自监督学习方法在高光谱图像分类领域已经取得了很大的进展,尽管如此,仅用少量的标记样本对高光谱遥感影像数据进行分类仍然存在一定的困难.本文提出了基于深度学习的 HSI 小样本分类算法和基于关系网络的 HSI 小样本分类算法,为解决小样本高光谱图像分类问题提供了新的方法.受到嵌入学习算法的启发,本文首先使用基于深度三维神经网络的嵌入模型从训练数据集中学习投影函数,学习到的投影函数将视觉特征映射到语义表示,并通过最小化损失函数使其具有较小的类内间距和较大的类间间距;再使用预先训练好的网络将每个未标记的目标类表示在相同的嵌入空间中;最后分别选用欧氏距离以及卷积神经网络作为距离度量方式来计算嵌入空间中嵌入与类原型之间的距离,从而对每个投影后的样本进行分类.本文在三个广泛使用的 HSI 数据集上进行了实验,结果表明所提的两种算法与基线方法 DFSL 算法相比均取得了更为理想的分类结果.

尽管算法取得了不错的表现,但仍需要更多研究来进一步提高算法的性能.在未来的工作中,我们将尝试利用高光谱隐含于光谱波段中的物理意义以提高算法的分类准确率.

参 考 文 献

- [1] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines[J]. *IEEE Trans Geosci Remote Sens*, 2004, 42(8): 1778-1790.
- [2] Gislason P O, Benediktsson J A, J R Sveinsson. Random forests for land cover classification[J]. *Pattern Recognit Lett*, 2006, 27(4): 294-300.
- [3] Chavez Jr P S, Kwarteng A Y. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis[J]. *Photogrammetric Engineering and Remote Sens*, 1989, 55(3): 1-6.
- [4] Villa A, Benediktsson J A, J Chanussot, et al. Hyperspectral image classification with independent component discriminant analysis[J]. *IEEE Trans Geosci Remote Sens* Dec, 2011, 49(12): 4865-4876.
- [5] Cai Y, Liu X, Cai Z. BS-Nets: an end-to-end framework for band selection of hyperspectral image [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 1: 1-16.
- [6] Chen Y, Lin Z, Zhao X, et al. Deep learning-based classification of hyperspectral data[J]. *IEEE J Sel Topics Appl, Earth Observ, Remote Sens*, 2014, 7(6): 2094-2107.
- [7] Chen Y, Zhao X, Jia X. Spectral-spatial classification of hyperspectral data based on deep belief network[J]. *IEEE J Sel Topics Appl, Earth Observ, Remote Sens*, 2015, 8(6): 2381-2392.
- [8] Mou L, Ghamisi P, Zhu X X. Deep recurrent neural networks for hyperspectral image classification[J]. *IEEE Trans Geosci Remote Sens*, 2017, 55, (7): 3639-3655.
- [9] Mei S, Ji J, Hou J, et al. Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks[J]. *IEEE Trans, Geosci Remote Sens*, 2017, 55(8): 4520-4533.
- [10] Rosenberg C, Hebert M, Schneiderman H. Semi-supervised selftraining of object detection models[J]. *IEEE Workshops Appl, Comput, Vis* Jan, 2005, 1: 29-36.
- [11] Ando R K, Zhang T. Two-view feature generation model for semi-supervised learning[J]. *24th Int Conf Mach Learn*, 2007, 1: 25-32.
- [12] Wang L, Hao S, Wang Q, et al. Semi-supervised classification for hyperspectral imagery based on spatial-spectral label propagation[J]. *ISPRS J Photogram Remote Sens*, 2014, 11(97): 123-137.
- [13] Kemker R, Kanan C. Self-taught feature learning for hyperspectral image classification[J]. *IEEE Trans, Geosci Remote Sens* May, 2017, 55(5): 2693-2705.
- [14] Dópido I, Li J, Marpu P R, et al. Semisupervised self-learning for hyperspectral image classification [J]. *IEEE Trans Geosci Remote Sens*, 2013, 51(7): 4032-4044.
- [15] Li W, Wu G, Zhang F, et al. Hyperspectral image classification using deep pixel-pair features[J]. *IEEE Trans Geosci Remote Sens*, 2017, 55(2): 844-853.
- [16] Pan B, Shi Z, Xu X. MugNet: Deep learning for hyperspectral image classification using limited samples[J]. *ISPRS J Photogrammetry Remote Sens*, 2018, 145: 108-119.
- [17] Zhan Y, Hu D, Wang Y, et al. Semisupervised hyperspectral image classification based on generative adversarial networks[J]. *IEEE Geosci Remote Sens Lett*, 2018, 15(2): 212-216.
- [18] Santoro A, Bartunov S, Botvinick M, et al. One-shot learning with memory-augmented neural networks[J]. *ArXiv Preprint ArXiv*, 2016, 1: 06065.
- [19] Finn, Chelsea, Pieter Abbeel, et al. Model-agnostic meta-learning for fast adaptation of deep networks[J]. *Proceedings of the 34th Interna-*

- tional Conference on Machine Learning, JMLR org, 2017, 70:1126-1135.
- [20] Oriol Vinyals, Charles Blundell, Tim Lillicrap, et al. Matching networks for one shot learning[J]. In Advances in Neural Information Processing Systems, 2016, 16:3630-3638.
- [21] Snell, Jake, Kevin Swersky, et al. Prototypical networks for few-shot learning[J]. Advances in Neural Information Processing Systems, 2017, 1:1-6.
- [22] Sung F, Yang Y, Zhang L, et al. Learning to Compare: Relation Network for Few-shot Learning[C]. //IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
- [23] Schwartz E, Karlinsky L, Shtok J, et al. An effective sample synthesis method for few-shot object recognition[J]. Advances in Neural Information Processing Systems Deltaencoder, 2018, 18:2850-2860.
- [24] Kwitt R, Hegenbart S, Niethammer M. One-shot learning of scene locations via feature trajectory transfer[J]. Conference on Computer Vision and Pattern Recognition, 2016, 16:78-86.
- [25] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features[J]. International Conference on Computer Vision, 2017, 17:1-7.
- [26] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. In Advances in Neural Information Processing Systems, 2014, 14:2672-2680.
- [27] Liu B, Yu X, Zhang P, et al. Supervised deep feature extraction for hyperspectral image classification[J]. IEEE Trans Geosci Remote Sens, 2018, 56(4):1909-1921.
- [28] Liu B, Yu X, Yu A, et al. Deep few-shot learning for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(4):2290-2304.
- [29] Misra, Diganta. Mish: A Self Regularized Non-Monotonic Neural Activation Function[J]. ArXiv August, 2019, 1908:08681.
- [30] Guangyong C, Pengfei C, Yujun S, et al. Rethinking the Usage of Batch Normalization and Dropout in the Training of Deep Neural Networks[J]. ArXiv May, 2019, 1905:05928.
- [31] Yokoya N, Iwasaki A. Airborne hyperspectral data over Chikusei[J]. Space Appl Lab, Univ Tokyo, Japan, 2016, 5:1-12.

HSI Classification Algorithm Based on Few-Shot Learning

ZHANG Jing YUAN Xi-guo

(School of Computer Science and Technology, Xidian University, Xi'an 710026, China)

Abstract Hyperspectral Image (HSI) contain rich spatial and spectral information, which provides favorable conditions to identify different classes of terrain accurately. At present, the performance of hyperspectral image classification method is relatively undesired with few labeled samples. Therefore, it is an important and challenging problem to give full play to the relevant theories and methods in few-shot learning to improve the classification performance of hyperspectral image classifiers. For this reason, based on few samples labeled, this paper improved the existing hyperspectral image classification algorithm based on deep learning. Above all, an embedded model based on the deep 3D-CNN is used to learn the projection function from the training datasets, which maps the visual features to the semantic representation, and make samples from the same class close and those from different classes far by minimizing the loss function. Then the pretrained network is employed to represent each unlabeled samples in the same embedding space. Finally, the Euclidean distance and CNN are respectively used as distance metric to obtain the distance between the embedding feature and the class prototype, so as to classify the samples in the test datasets. Experiments are carried out on the Indian pines, Salinas and Pavia University, and the experimental results show that both the improved algorithm based on deep learning and algorithm based on relation network can achieve the desired classification performance.

Key words few-shot learning; hyperspectral image; HSI classification; deep learning